Rationality & Paradox, Part I:

The Liar;
The Barber;
(The Knowability Paradox);
& a Real-Life Paradox at Altitude …

Selmer Bringsjord

Are Humans Rational?

10/21/19

RAIR
Rensselaer AI and Reasoning Lab

Rationality & Paradox, Part I:
The Liar;
The Barber;
(The Knowability Paradox);
& a Real-Life Paradox at Altitude …

Selmer Bringsjord

Are Humans Rational?

10/21/19

R A I R
Rensselaer AI and Reasoning Lab

Rationality & Paradox, Part I:

The Liar;
The Barber;
(The Knowability Paradox);
& a Real-Life Paradox at Altitude …

Selmer Bringsjord

Are Humans Rational?

10/21/19

Rensselaer AI and Reasoning Lab

In our context,
why study paradoxes?

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek*, *etc*).

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek, etc*).
- Because solving paradoxes/apparent contradictions can light the path forward in logic, mathematics, AI, etc. — & can even save lives!

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek*, *etc*).

Mike

- Because solving paradoxes/apparent contradictions can light the path forward in logic, mathematics, AI, etc. — & can even save lives!

Types of Paradoxes

- Deductive Paradoxes — corresponding to Area 1, going all the way back to our syllabus
- Inductive Paradoxes — corresponding to Area 2

Types of Paradoxes

Today; Part I (& then Oct 28, Oct 31, Nov

- Deductive Paradoxes — corresponding to Area 1, going all the way back to our syllabus
- Inductive Paradoxes — corresponding to Area 2

The Liar (Paradox) …

First, …
remember/know:

First, …
remember/know:

p iff $\neg p$ (for any statement p)

First, …
remember/know:

p iff $\neg p$ (for any statement p)

is a contradiction: it's equivalent to p & $\neg p$.

First, …
remember/know:

p iff $\neg p$ (for any statement p)

is a contradiction: it's equivalent to p & $\neg p$.

(Why?)

The (Economical) Liar

The (Economical) Liar

L: This sentence is false.

The (Economical) Liar

L: This sentence is false.

If T(**L**) then ¬T(**L**)

The (Economical) Liar

L: This sentence is false.

If $T(\mathbf{L})$ then $\neg T(\mathbf{L})$

If $\neg T(\mathbf{L})$ then $T(\mathbf{L})$

The (Economical) Liar

L: This sentence is false.

If T(**L**) then ¬T(**L**)

If ¬T(**L**) then T(**L**)

T(**L**) iff (i.e., if & only if) ¬T(**L**)

The (Economical) Liar

L: This sentence is false.

If $T(\mathbf{L})$ then $\neg T(\mathbf{L})$

If $\neg T(\mathbf{L})$ then $T(\mathbf{L})$

$T(\mathbf{L})$ iff (i.e., if & only if) $\neg T(\mathbf{L})$

Contradiction!

The “Gödelian” Liar

The "Gödelian" Liar

$\bar{P}$: This sentence is unprovable.

The “Gödelian” Liar

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable (since from the supposition that some statement S holds we can prove $S \rightarrow S$ and then promptly deduce S by *modus ponens*). But since what $\bar{P}$ says is that it's unprovable, we have that $\bar{P}$ is false under our supposition.

The “Gödelian” Liar

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable (since from the supposition that some statement S holds we can prove $S \to S$ and then promptly deduce S by *modus ponens*). But since what $\bar{P}$ says is that it’s unprovable, we have that $\bar{P}$ is false under our supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable (since if $\bar{P}$ were provable it would be true, because what it says is that it’s unprovable and we would have proved what it says). But since what $\bar{P}$ says is that it’s unprovable, and we proved $\bar{P}$ under our supposition, $\bar{P}$ is true.

The "Gödelian" Liar

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable (since from the supposition that some statement S holds we can prove $S \rightarrow S$ and then promptly deduce S by *modus ponens*). But since what $\bar{P}$ says is that it's unprovable, we have that $\bar{P}$ is false under our supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable (since if $\bar{P}$ were provable it would be true, because what it says is that it's unprovable and we would have proved what it says). But since what $\bar{P}$ says is that it's unprovable, and we proved $\bar{P}$ under our supposition, $\bar{P}$ is true.

$\mathrm{T}(\bar{P})$ iff (i.e., if & only if) $\neg\mathrm{T}(\bar{P}) = \mathrm{F}(\bar{P})$

The "Gödelian" Liar

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable (since from the supposition that some statement S holds we can prove $S \rightarrow S$ and then promptly deduce S by *modus ponens*). But since what $\bar{P}$ says is that it's unprovable, we have that $\bar{P}$ is false under our supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable (since if $\bar{P}$ were provable it would be true, because what it says is that it's unprovable and we would have proved what it says). But since what $\bar{P}$ says is that it's unprovable, and we proved $\bar{P}$ under our supposition, $\bar{P}$ is true.

T($\bar{P}$) iff (i.e., if & only if) ¬T($\bar{P}$) = F($\bar{P}$)

Contradiction!

The (Verbose) Liar — With a Twist

The (Verbose) Liar — With a Twist

Theorem: $2+2 = 5$.

The (Verbose) Liar — With a Twist

Theorem: 2+2 = 5.

Proof: Set:

The (Verbose) Liar — With a Twist

Theorem: 2+2 = 5.

Proof: Set:

L: This sentence is false.

The (Verbose) Liar — With a Twist

Theorem: 2+2 = 5.

Proof: Set:

L: This sentence is false.

L is either true or false. Suppose that it's true. Then since what it says is that it's false, it *is* false; i.e., **L** is false, on this supposition. So we've proved that if **L** is true, L is false. Now suppose instead that **L** is false. Then since it says that it's false, it's true; i.e., **L** is true, on our current supposition. We have thus proved that if **L** is false, **L** is true. Combining the conditionals we've proved yields this: **L** is true if and only if **L** is false, which is a contradiction. (*P* if and only if ¬*P* is logically equivalent to *P* and ¬*P*.) By the rule of inference *explosion*, it follows that 2+2 = 5. **QED**

Outlawing Self-Referential Sentences Isn't the Answer!

Outlawing Self-Referential Sentences Isn't the Answer!

- For the following sentences, e.g., are perfectly innocuous, and obviously true or false (only) as the case may be, without complication.

Outlawing Self-Referential Sentences Isn't the Answer!

- For the following sentences, e.g., are perfectly innocuous, and obviously true or false (only) as the case may be, without complication.
  - `This sentence is a sentence.`

Outlawing Self-Referential Sentences Isn't the Answer!

- For the following sentences, e.g., are perfectly innocuous, and obviously true or false (only) as the case may be, without complication.
  - `This sentence is a sentence.`
  - `This sentence contains the letter ‘r’.`

Outlawing Self-Referential Sentences Isn't the Answer!

- For the following sentences, e.g., are perfectly innocuous, and obviously true or false (only) as the case may be, without complication.
  - `This sentence is a sentence.`
  - `This sentence contains the letter 'r'.`
  - `This sentence has more than three letters in it.`

Outlawing Self-Referential Sentences Isn't the Answer!

- For the following sentences, e.g., are perfectly innocuous, and obviously true or false (only) as the case may be, without complication.
  - `This sentence is a sentence.`
  - `This sentence contains the letter ‘r’.`
  - `This sentence has more than three letters in it.`
  - `This sentence ends with a period, starts with a capital ’T’,  and has more than two words.`

Outlawing Self-Referential Sentences Isn't the Answer!

- For the following sentences, e.g., are perfectly innocuous, and obviously true or false (only) as the case may be, without complication.
  - `This sentence is a sentence.`
  - `This sentence contains the letter ‘r’.`
  - `This sentence has more than three letters in it.`
  - `This sentence ends with a period, starts with a capital ’T’,  and has more than two words.`
  - ...

Outlawing Self-Referential Sentences Isn't the Answer!

Outlawing Self-Referential Sentences Isn't the Answer!

Box 1

Outlawing Self-Referential Sentences Isn't the Answer!

Box 1

Box 2

Outlawing Self-Referential Sentences Isn't the Answer!

Box 1	Box 2
The sentence in Box 2 is true.	The sentence in Box 1 is false.

Outlawing Self-Referential Sentences Isn't the Answer!

Box 1		Box 2
The sentence in Box 2 is true.	Neither sentence is self-referential.	The sentence in Box 1 is false.

Outlawing Self-Referential Sentences Isn't the Answer!

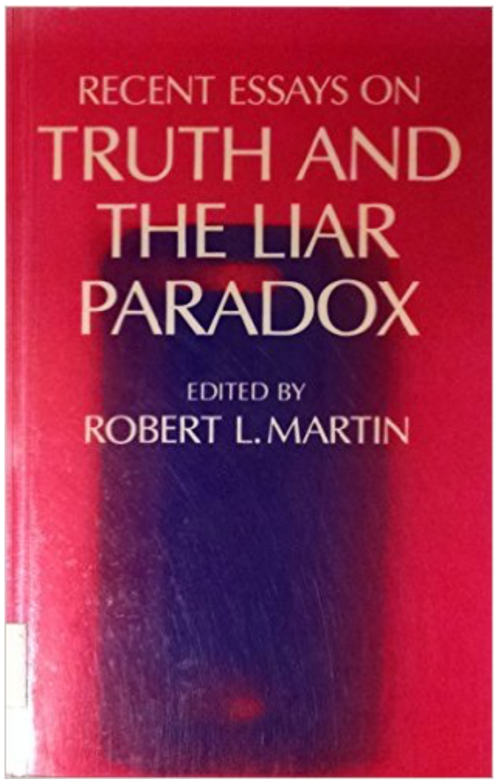
Box 1		Box 2
The sentence in Box 2 is true.	Neither sentence is self-referential.	The sentence in Box 1 is false.

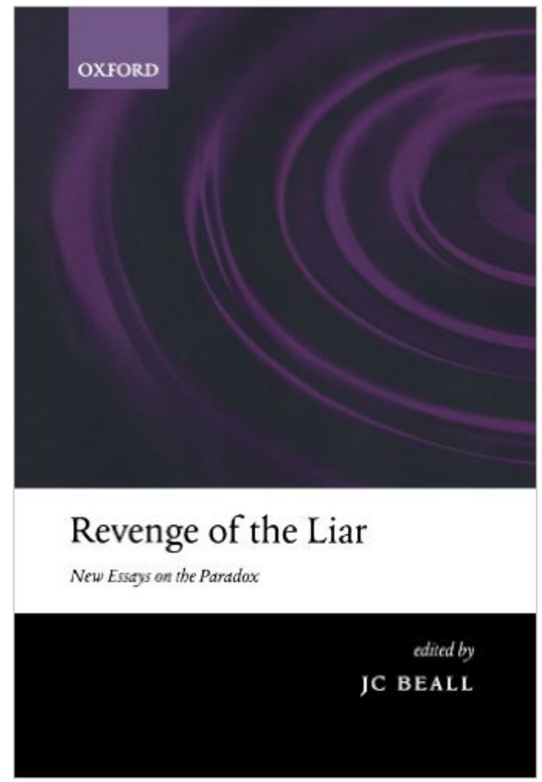
Outlawing Self-Referential Sentences Isn't the Answer!

Box 1		Box 2
The sentence in Box 2 is true.	Neither sentence is self-referential.	The sentence in Box 1 is false.

Suppose that the sentence in Box 1 is true. Then the sentence in Box 2 is true (because the sentence in Box 1 says that that sentence is true). But then the sentence in Box 1 is false (because the sentence in Box 2 says that that sentence is false). So, if the sentence in Box 1 is true, it's false. On the other hand, if the sentence in Box 1 is false, the sentence in Box 2 is false. But then the sentence in Box 1 is true; so we've shown that if the sentence in Box 1 is false, it's true. We thus have again a contradiction: The sentence in Box 1 is true if and only if it's not true!

Further Reading …





The Barber Paradox …

S1: There was once a small town in Norway in which there was a barber who shaved all and only the men residing in the town who didn't shave themselves.

S1: There was once a small town in Norway in which there was a barber who shaved all and only the men residing in the town who didn't shave themselves.

$\vdash \phi \wedge \neg\phi$?

S1: There was once a small town in Norway in which there was a barber who shaved all and only the men residing in the town who didn't shave themselves.

$\vdash \phi \wedge \neg\phi?$

S2: There was once a small town in Norway in which there resided a barber who shaved all and only the men residing in the town who didn't shave themselves.

S1: There was once a small town in Norway in which there was a barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi$?

S2: There was once a small town in Norway in which there resided a barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi$?

S1: There was once a small town in Norway in which there was a barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi?$

S2: There was once a small town in Norway in which there resided a barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi?$

S3: There was once a small town in Norway in which there resided a male barber who shaved all and only the men residing in the town who didn't shave themselves.

S1: There was once a small town in Norway in which there was a barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi$?

S2: There was once a small town in Norway in which there resided a barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi$?

S3: There was once a small town in Norway in which there resided a male barber who shaved all and only the men residing in the town who didn't shave themselves. $\vdash \phi \wedge \neg\phi$?

Situation S3 is impossible!

Situation S3 is impossible!

Proof: Let's assume for the sake of argument that such a situation can be. Without loss of generality, let the town be Lyngdal and the male Lyngdalian barber be Olaf. Either Olaf shaves himself or he doesn't. But either case leads straight to a contradiction. Therefore the situation is in fact impossible. Here we go …

Suppose Olaf shaves himself. Then it follows that he doesn't shave himself. Suppose on the other hand that Olaf doesn't shave himself. Then is follows that he does shave himself. Hence, Olaf shaves himself if and only if he doesn't shave himself, which is a contradiction. **QED**

For homework …

For Presently Rational System-2ers (1/2)

- **Argument**: (1) Harrison admires only great actors who don't admire themselves. (2) Harrison admires all great actors who don't admire themselves. Therefore, (3) Harrison isn't a great actor.
- **Purported Proof to Certify the Argument**: For indirect proof, suppose that H is a great actor. Either H admires himself, or he doesn't. We show that either case leads to a contradiction, and hence that our starting supposition is false, i.e. H isn't a great actor. First, suppose that H does admire himself. From this and (1) we deduce that H doesn't admire himself — contradiction. For the second case, suppose that H doesn't admire himself. From this and (2) it follows that Harrison does admire himself — contradiction. Either alternative needs to the desired contradiction. **QED**
- Is the purported proof in fact a *valid* proof?

From Language, Proof, and Logic (2000) by Barwise & Etchemendy, p. 335.

For homework …

For Presently Rational System-2ers (2/2)

Show that it would be impossible to construct a reference book that lists all and only those reference books that do not list themselves.

From Language, Proof, and Logic (2000) by Barwise & Etchemendy, p. 337.

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek*, *etc*).
- Because solving paradoxes/apparent contradictions can light the path forward in logic, mathematics, AI, etc. — & can even save lives!

In our context, why study paradoxes?

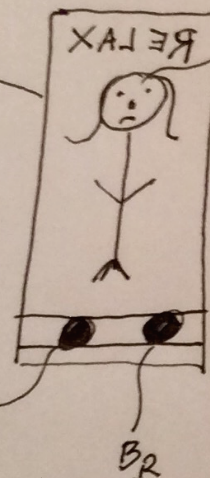
- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek, etc*).
- Because solving paradoxes/apparent contradictions can light the path forward in logic, mathematics, AI, etc. — & can even save lives!

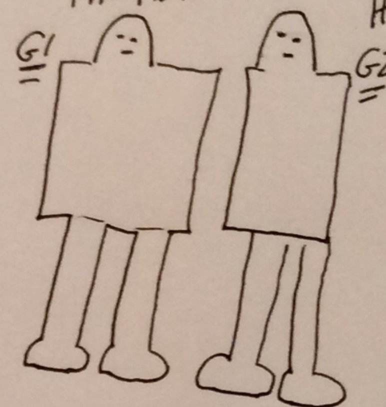
Dr Who and …
The Problem of the Osirians …

THE PROBLEM OF THE OSIRIANS

ONE QUESTION ALLOWED!

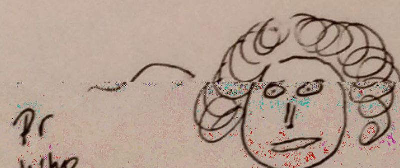
The Twin Guardians of Horus





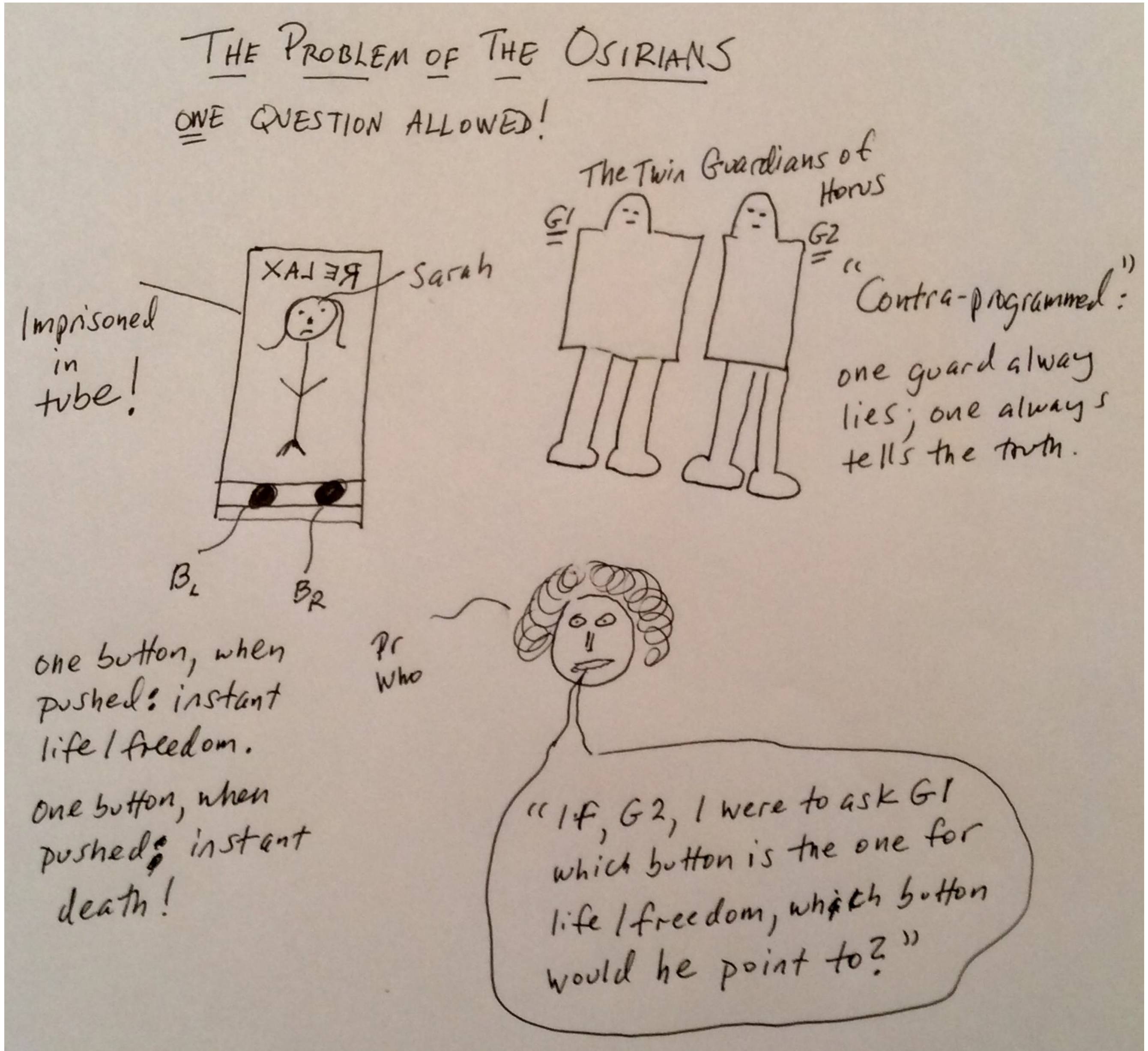
"Contra-programmed":

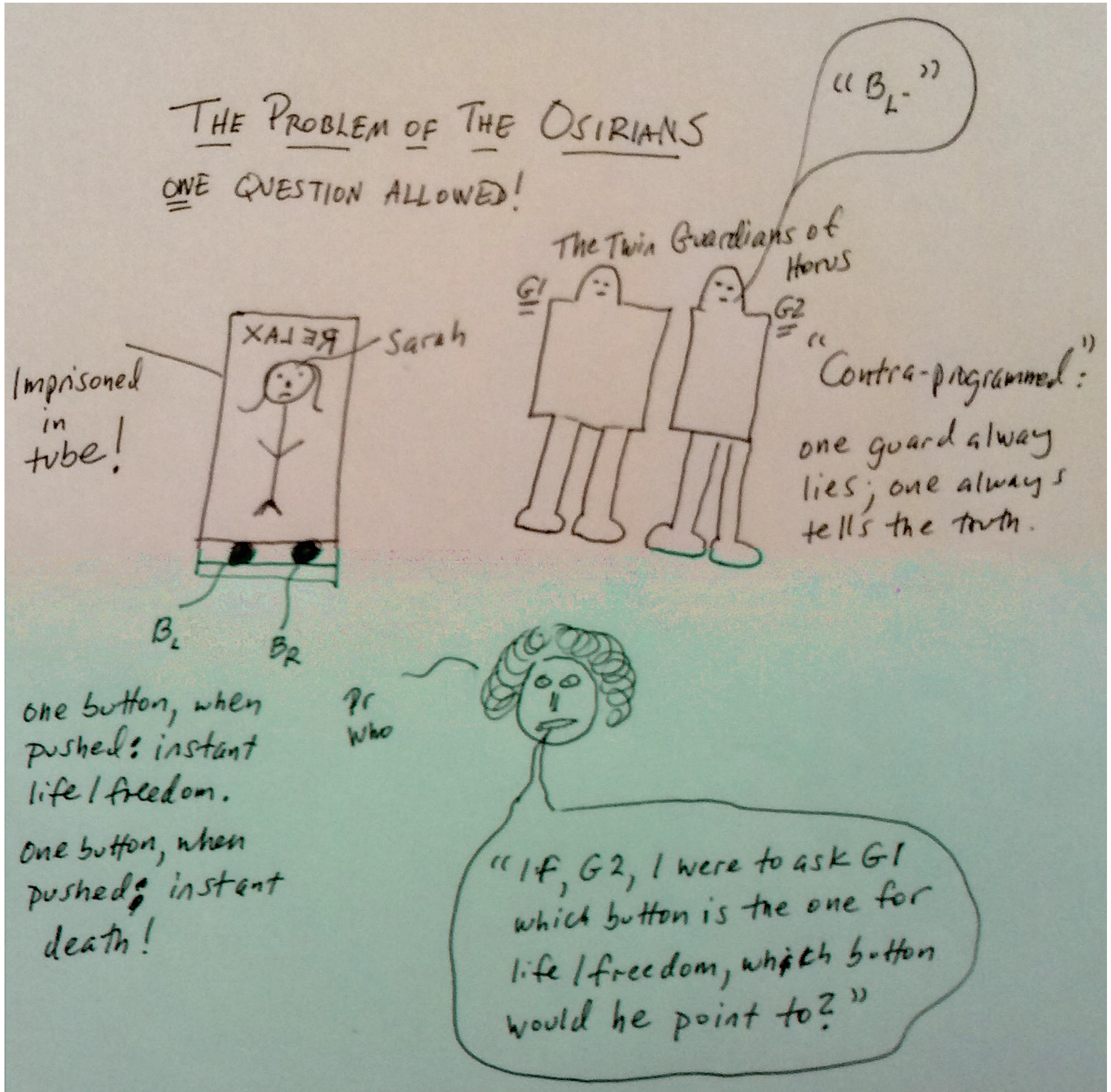
one guard always lies; one always tells the truth.



one button, when pushed: instant life/freedom.

one button, when pushed: instant death!


THE PROBLEM OF THE OSIRIANS
ONE QUESTION ALLOWED!
The Twin Guardians of Horus
G1
G2
"Contra-programmed":
one guard alway lies; one always tells the truth.
RELAX
Sarah
Imprisoned in tube!
B_L
B_R
Dr Who
one button, when pushed: instant life/freedom.
One button, when pushed: instant death!
"If, G2, I were to ask G1 which button is the one for life/freedom, which button would he point to?"


THE PROBLEM OF THE OSIRIANS
ONE QUESTION ALLOWED!
"B_L."
The Twin Guardians of Horus
G1
G2
"Contra-programmed":
one guard alway lies; one always tells the truth.
RELAX
Sarah
Imprisoned in tube!
B_L
B_R
Dr Who
one button, when pushed: instant life/freedom.
One button, when pushed: instant death!
"If, G2, I were to ask G1 which button is the one for life/freedom, which button would he point to?"

Dr Who's (Background) Liar-Leveraging Proof

Proposition: B_R leads to freedom and life!

Proof: G2, who has uttered "B_L," is either a liar or a truth-teller. Suppose, first, the former case. Then G1 would in fact say "B_R" and would be telling the truth in so saying; hence B_R on this first supposition is the way to go. What about the latter case? In this case, G2 is telling the truth, and hence G1 would in fact say "B_L" — but would be lying, so in this case B_R is again the way to go. **QED**

Gödel proved his first incompleteness theorem, one of the greatest achievements in the history of mathematics, by using reasoning that parallels the reasoning in The Liar!!

Gödel proved his first incompleteness theorem, one of the greatest achievements in the history of mathematics, by using reasoning that parallels the reasoning in The Liar!!

Syllabus:Schedule:December 2!

Gödel proved his first incompleteness theorem, one of the greatest achievements in the history of mathematics, by using reasoning that parallels the reasoning in The Liar!!

Syllabus:Schedule:December 2!

Study the “Gödelian” Liar given earlier in this deck.

For Further (Supererogatory) Study: The Knowability Paradox …

Don't blame me. The true source only recently discovered in 2009: Alonzo Church, Turing's PhD advisor in the States, and the inventor of the lambda-calculus.

The Knowability Paradox (informal)

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \mathbin{\&} \neg\mathbf{K}p^*) \qquad (1)$$

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \mathbin{\&} \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \mathbin{\&} \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \;\&\; \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \;\&\; \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \;\&\; \neg\mathbf{K}p^* \qquad (3)$$

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (4)$$

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \mathbin{\&} \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \mathbin{\&} \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \mathbin{\&} \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \mathbin{\&} \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

$$\text{Necessarily: } \neg\mathbf{K}(p^* \mathbin{\&} \neg\mathbf{K}p^*) \qquad (5)$$

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

$$\text{Necessarily: } \neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (5)$$

And in a final move, we know that from 'necessarily it's not the case that' it follows that 'it's not possible that it's the case that', so we can deduce from (5):

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

$$\text{Necessarily: } \neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (5)$$

And in a final move, we know that from 'necessarily it's not the case that' it follows that 'it's not possible that it's the case that', so we can deduce from (5):

$$\text{It's not possible that: } \mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (6)$$

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \ \& \ \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \ \& \ \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then *s* holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \ \& \ \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \ \& \ \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

$$\text{Necessarily: } \neg\mathbf{K}(p^* \ \& \ \neg\mathbf{K}p^*) \qquad (5)$$

And in a final move, we know that from 'necessarily it's not the case that' it follows that 'it's not possible that it's the case that', so we can deduce from (5):

$$\text{It's not possible that: } \mathbf{K}(p^* \ \& \ \neg\mathbf{K}p^*) \qquad (6)$$

What (6) says is this: There is an *unknowable* truth! — for what (6) declares is this: "It's impossible that someone at some time knows that the conjunction of p* & ¬Kp* holds." !

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

Letting **K** represent 'someone at some time knows,' our supposition can be conveniently represented as the statement:

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then s holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

$$\text{Necessarily: } \neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (5)$$

And in a final move, we know that from 'necessarily it's not the case that' it follows that 'it's not possible that it's the case that', so we can deduce from (5):

$$\text{It's not possible that: } \mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (6)$$

What (6) says is this: There is an *unknowable* truth! — for what (6) declares is this: "It's impossible that someone at some time knows that the conjunction of p* & ¬Kp* holds." !

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?) Letting **K** represent 'someone at some time knows', our supposition can be semi-formally represented as the statement:

Why is this (at least according to some) a paradox??

$$\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (1)$$

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that someone at some time knows that snow is white, and someone at some time knows that Tromsø is in Norway. Hence:

$$\mathbf{K}p^* \,\&\, \mathbf{K}\neg\mathbf{K}p^* \qquad (2)$$

Next, if someone at some time knows a proposition *s*, then s holds. We can apply this to the right conjunct in (2) to obtain:

$$\mathbf{K}p^* \,\&\, \neg\mathbf{K}p^* \qquad (3)$$

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition that led to the contradiction. I.e., we deduce:

$$\neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (4)$$

We're not done yet, not by a long shot! For (4) is an outright theorem, and hence holds necessarily; this yields:

$$\text{Necessarily: } \neg\mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (5)$$

And in a final move, we know that from 'necessarily it's not the case that' it follows that 'it's not possible that it's the case that', so we can deduce from (5):

$$\text{It's not possible that: } \mathbf{K}(p^* \,\&\, \neg\mathbf{K}p^*) \qquad (6)$$

What (6) says is this: There is an *unknowable* truth! — for what (6) declares is this: "It's impossible that someone at some time knows that the conjunction of p* & ¬Kp* holds." !

Why is this (at least according to some) a paradox??

Because it's supposed to be absurd that sitting in our armchairs, starting from the seemingly innocent and obvious fact that — and this is represented by (1) — there's at least one unknown truth, we can deduce that — and this is (6) — there's a truth that is absolutely unknow*able*.

The Knowability Paradox (informal)

Suppose that we know that there is at least one unknown truth; let's call it p*; this is an entirely arbitrary name. (While this is a supposition, it would be a very plausible assertion. After all, we're finite, non-omniscient beings. And don't we have examples?)

represented as the statement:

K(p* & ¬**K**p*) (1)

Obviously, knowledge distributes over conjunctions. After all, if someone at some time knows that snow is white and that Tromsø is in Norway, it logically follows that

(2)

the right conjunct in (2) to obtain:

But (3) is a contradiction. Hence by *reductio ad absurdum*, we reject the original supposition

(4)

Necessarily: ¬**K**(p* & ¬**K**p*) (5)

And in a final move, we know that from 'necessarily it's not the case that' it follows that 'it's not possible that it's the case that', so we can deduce from (5):

It's not possible that: **K**(p* & ¬**K**p*) (6)

What (6) says is this: There is an *unknowable* truth! — for what (6) declares is this: "It's impossible that someone at some time knows that the conjunction of p* & ¬Kp* holds." !

Invalid?

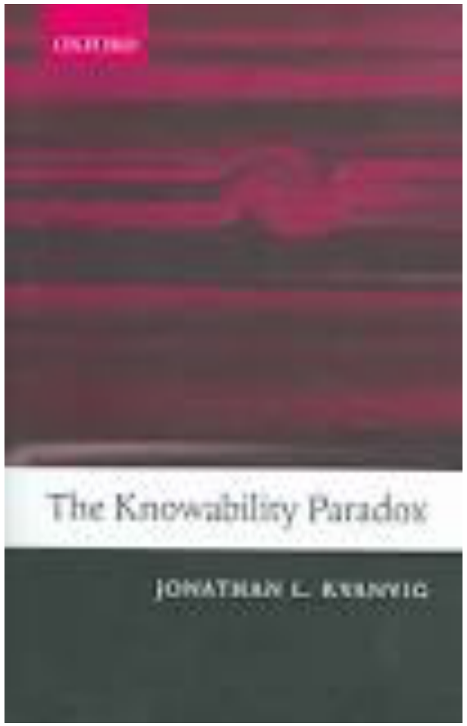
Invalid?

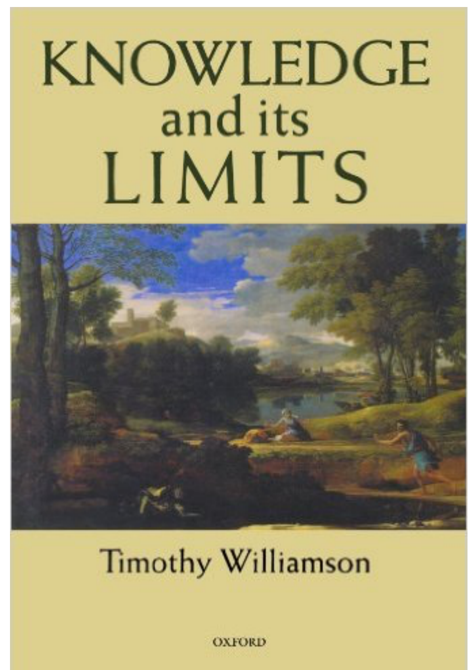
- John Andrews, Lauren Talbot, Hannah F., Christina Elmore were concerned that the initial statement isn't something like: "There exists some proposition *p* which is such that, *p* holds, and no one knows that *p*."

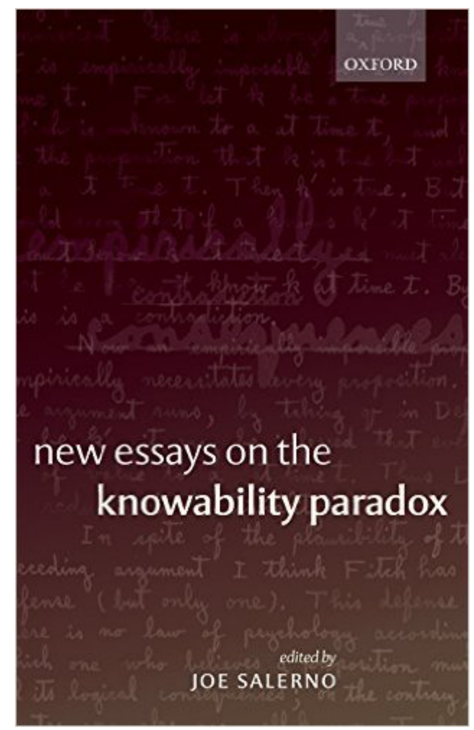
Invalid?

- John Andrews, Lauren Talbot, Hannah F., Christina Elmore were concerned that the initial statement isn't something like: "There exists some proposition *p* which is such that, *p* holds, and no one knows that *p*."
  - Both Kvanvig and Williamson say that if you start with this, you get a parallel deduction that confirms the statement that there's an absolutely unknowable truth.

Further Reading …







```
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

{:name        "Knowability paradox"
 :description " exists p  ~Diamond exists x Kx (Tp & ~ exist y Ky Tp)"

 :assumptions {}
 :goal (exists [?P] (not (pos (exists [?x] (Knows! ?x (and ?P (not (exists [?y] (Knows! ?y ?P))))))))))}


;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
```

Sandbox

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_112.jdk/Contents/Home/bin/java ...
```

```
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

{:name        "Knowability paradox"
 :description " exists p  ~Diamond exists x Kx (Tp & ~ exist y Ky Tp)"

 :assumptions {}
 :goal (exists [?P] (not (pos (exists [?x] (Knows! ?x (and ?P (not (exists [?y] (Knows! ?y ?P))))))))}


;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
```

Sandbox

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_112.jdk/Contents/Home/bin/java ...
```

```
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

{:name        "Knowability paradox"
 :description " exists p  ~Diamond exists x Kx (Tp & ~ exist y Ky Tp)"

 :assumptions {}
 :goal (exists [?P] (not (pos (exists [?x] (Knows! ?x (and ?P (not (exists [?y] (Knows! ?y ?P))))))))}

;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
```

Sandbox

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_112.jdk/Contents/Home/bin/java ...
```

$$\exists \phi \neg \Diamond \exists a \mathbf{K}[a, T(\phi) \wedge \neg \exists a' \mathbf{K}(a', T(\phi))]$$

In our context,
why study paradoxes?

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek*, *etc*).

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek*, *etc*).
- Because solving paradoxes/apparent contradictions can light the path forward in logic, mathematics, AI, etc. — & can even save lives!

In our context, why study paradoxes?

- Because anyone defending ***R*** had better be able to resolve paradoxes!
- Because such study requires and develops System-2 cognition, quite possibly better than any other study.
- Because intellectual humility in humans is a good thing (and intellectual overconfidence is a bad thing), and such study is a road to intellectual humility (and an antidote to intellectual overconfidence).
- Because it can save the day, sometimes (Dr Who, *Star Trek*, *etc*).

Mike

- Because solving paradoxes/apparent contradictions can light the path forward in logic, mathematics, AI, etc. — & can even save lives!

slutten